On the ethnic classification of Pakistani face using deep learning

Jilani, Shelina Khalid; Ugail, Hassan; Bukar, Ali Maina; Logan, Andrew

Published in:
2019 International Conference on Cyberworlds (CW)

DOI:
10.1109/CW.2019.00039

Publication date:
2019

Document Version
Author accepted manuscript

On the Ethnic Classification of Pakistani Face using Deep Learning

Shelina Khalid Jilani*, Hassan Ugail, A. M. Bukar.
Centre for Visual Computing
University of Bradford
Bradford, UK
S.Jilani@bradford.ac.uk

Andrew Logan
Department of Vision Science
Glasgow Caledonian University
Glasgow, Scotland.

Abstract—

Demographic-based identification plays an active role in the field of face identification. Over the past decade, machine learning algorithms have been used to investigate challenges surrouding ethnic classification for specific populations, such as African, Asian and Caucasian people. Ethnic classification for individuals of South Asian, Pakistani heritage, however, remains to be addressed.

The present paper addresses a two-category (Pakistani Vs Non-Pakistani) classification task from a novel, purpose-built dataset. To the best of our knowledge, this work is the first to report a machine learning ethnic classification task with South Asian (Pakistani) faces. We conduted a series of experiments using deep learning algorithms (ResNet-50, ResNet-101 and ResNet-152) for feature extraction and a linear support vector machine (SVM) for classification. The experimental results demonstrate ResNet-101 achieves the highest performance accuracy of 99.2% for full-face ethnicity classification, followed closely by 91.7% and 95.7% for the nose and mouth respectively.

Keywords— Ethnicity; Pakistani; Deep Learning; Residual Network; Classification.

I. INTRODUCTION

For humans, a brief glimpse of a face can be sufficient to reveal details of an individual's age, gender, ethnicity, emotional state and direction of attention. A visual attribute of the face, ethnicity is perhaps the least studied soft biometric within literature, possibly because the physical traits of human populations remain undefined. Further within populations people can exhibit certain characteristics to a greater or lesser degree, generating intra-ethnic variations in physiology [1]. As a result, the application of ethnicity as a soft biometric trait for facial image classification remains a long-standing and challenging task within the field of Machine Learning.

Nevertheless, there is a distinction between facial features of different countries [2]. For example, a Chinese, Japanese and a Korean face may appear visually alike to people who are not familiar with this region, possibly due to the proximity between the countries. But differences between the soft-tissue profile of Korean and European-Americans for example, are reported and include a host of differences, including sexual dimorphism [3].

Likewise, faces of Indian and Pakistani origin may also superficially appear similar. However, unlike Bangladesh for example, where the ethnicity is largely homogeneous [4]. People of Pakistani origin can belong to one of the many diverse ethnic and ethno-linguistic groups such as Baloch, Kashmiris, Punjabi, Pashtun, Sindhi and Mujahirs [5] [6]. The diversity of the Pakistani human groups reflects individual regions of the country, which in turn creates a spectrum of different facial appearances.

The most comprehensive review on learning race from a face was written by Fu et al., [7]. By using state-of-the-art methods, the researchers proposed two ways race classification could be achieved; (1) as a single modality or (2) with the incorporation of multi-modals methods, such as the fusion of face and gait, for example.

In recent times, Deep Learning [8], which functions by automatically learning the best representation of features from images, has attracted significant attention from researchers due to its multi-faceted application in computer vision [9] [10]. While race classification has been attempted it is limited to commonly known populations such as African-American, Caucasian and Asian populations. Ethnicity classification has been reported for the Chinese [11], Japanese [12] and Korean [13] face, however the South Asian, Pakistani group remains untried, with a single exception (Jilani et.al [14]).

In this work, we focus on ethnicity as a demographic trait, visible within the face, and conduct a series of experiments using pre-trained Deep Learning models called Residual Neural Network (ResNet). The models used as part of our proposed methodology are: ResNet-50, ResNet-101 and ResNet-152.

Our model is evaluated on a dataset of constrained criterion-specific, high resolution facial images. Due to the polysemy nature of the term ethnicity, and for the research presented here, we define ethnicity as a person's cultural and ancestral background. The method by which images were categorised as Pakistani and Non-Pakistani is discussed in the upcoming methodology section.

The proposed approach for ethnicity classification consists of 2 fundamental components; (i) Feature extraction using weights of ResNet-50, ResNet-101 and ResNet-152 and (ii) demographic (ethnicity) classification using Support Vector Machine (SVM) algorithm as a binary, two-class technique.

The paper is organized as follows; we present the concept of ethnicity in **section II**. In **section III**, we present literature related to the ethnic classification of facial images. A novel dataset created specifically for the experiments is presented as part of **section IV**. In **section V**, we discuss the methodology employed for the experiments. In **section VI** we present the experimental results. **Section VII** consists of a discussion and finally, we conclude our work in **section VIII**.

II. ETHNICITY & ANTHROPOMETRY

Anthropometric studies demonstrate that there are distinctions between feature measurements and characteristics of people from different racial and ethnic backgrounds. Leslie Farkas and colleagues carried out a comprehensive anthropometric study of facial morphology and facial parameters comparing 14 normative measurements of the face across ethnic and racial groups [15].

Several differences between groups were reported. Specifically, the orbital region and nose height showed the greatest discrepancies in measurements across all the researched groups (Africa, Asian, Europe, Middle-East and North Africa). In more detail, the nose was typically wide in both males and females of the Asian and African racial group. For the Middle Eastern group, however, the nose width was comparable to that of the North American Caucasians, but differed significantly in nasal height. The results of Farkas and colleagues (2005) were supported by a systematic review which investigated inter-ethnic variability of facial measurements [16].

Inter-ethnic variability was described by 95% confidence intervals of individual measurements. A Bayesian hierarchical random effects model was created to approximate posterior means and 95% credible intervals (CrI) for each measurement by ethnicity/race-group.

Fang et al. [16] showed that the forehead height (measured trichion (*tr*) – nasion (*n*) and the intercanthal distance (measured endocanthion (*en*) - endocanthion (*en*), demonstrated the greatest inter-ethic variation. Whereas, measurements of the mid-face width (measure zygomatic (zy) - zygomatic (zy) and the exocanthion distance (measured right ex - left ex) showed the lowest degree of variability.

III. RELATED LITERATURE

Deep learning methods have reported exceptional results on many image classification tasks using labelled training datasets. Researchers have worked on large scale databases such as ImageNet [17] to evaluate the performance of deep learning algorithms.

In 2015 Microsoft Research Asia developed Deep Residual Network (ResNet) [18]. The critical feature of the framework is its “identity shortcut connection” which functions to skip layers during learning without compromising accuracy. This in turn reduces computation time and increases performance accuracy.

Research conducted by Ou et. al [19] conducted a binary classification (Asian Vs Non-Asian class) using frontal face images. Real-time analysis was achieved using images from uncontrolled environments. Principal Component Analysis (PCA) was used to obtain the most variant features and a novel “321” algorithm was combined with a Support Vector Machine (SVM) to boost classification. The researchers reported an 82.5% classification accuracy on a database of 750 face images taken from The Facial Recognition Technology Database (FERET).

Hosoi et. al [2] investigated a three class (Asian Vs African Vs European) ethnicity classified task. From 1,991 facial images, key features were extracted by a combination of Gabor Wavelets Transformation (GWT) and retina sampling techniques. Since the classification was not binary, the Support Vector Machine used was expanded to a multi-clustering classifier. Based on the applied technique the estimation accuracy achieved was as follows: Asian: 96.3%, European: 93.1% and African: 94.3%.

Ahmed et.al [20] presented a framework of training a Convolutional Neural Network (CNN) model using transfer-learning from pseudo-tasks, to classify the ethnic origin of faces from the White, Asian and other race using the FERET and FRGC database.

Guo and Mu [21] used Biologically-Inspired Features (BIFs) with Manifold Learning and a Support Vector Machine classifier to estimate the ethnicity of Black, White, Hispanic, Asian and Indian face images. More recently, Han et. al [22] proposed an automatic multi-demographic (age, gender and race) framework. Biologically-Inspired Features (BIFs) were used to extract demographic informative features from facial images. And a Hierarchical classifier was applied to categorize face race.

Local Binary Patterns (LBP) and Weber Local Descriptors (WLD) were utilized by Muhammed et. al [23] for the racial classification of 5 distinct classes: Asian, Black, Hispanic, Middle-Eastern and White. A total of 1,188 images taken from the FERET database were used for training while 1,180 were used for the testing phase.

The performance accuracy achieved for LBP was 98.42%, 95.56%, 93.65%, 100% and 98.18% for the Asian, Black, Hispanic, Middle-Eastern and the White racial class respectively. Comparable results were achieved with WLD; 97.74%, 96.89%, 92.06%, 98.33% and 99.53% again for the Asian, Black, Hispanic, Middle-Eastern and the White race respectively.

In contrast, Tin and Sein, 2011 investigated a two-class racial classification task using only Myanmar and Non-Myanmar facial images [24]. A dataset of 250 images were collected from the internet and the results identified an average 94% accuracy, despite the small dataset.

More recently Lakshmiprabha [25] produced a multi-modal framework for gender, ethnicity, age and expression classification. Four different feature extraction methods were used and Principal Component Analysis (PCA) was used for feature dimensionality reduction while a Neural Network was used as a classifier. Images were collected from publically available datasets such as PAL, JAFFE and FERET. A total of 357 images from all three databases were collected and categorized into the following ethnic groups; White, Black, Indian and Other (consisted of Asian and Hispanic). Of the 357 images, 187 were used for testing and it was reported that the Active Appearance Model (AAM) generated the highest ethnic classification rate of 93.83%.

The literature on ethnic and racial classification features several consistent themes. Firstly, experiments have been conducted on previously published databases. This may be due to the limited demographic variability and restrictions which meant researchers used images harvested from the internet. The problem here is that we are presented with a range of results for various image-based classification problems, on unchanged datasets. Nonetheless it seems that the use of the FERET database remains the most common choice for researchers, possibly due to the size and variability of the dataset. However, the problem with using such an old database is that it is outdated in terms of the equipment used to capture the images, and may not reflect the most recent demographic trends.

Another observation is the complexity surrounding the definitions of the terms "race" and "ethnicity". The use of the words within the literature remains interchangeable but, where possible, researchers have drawn an association between facial appearance and geographical origin, i.e.: Chinese and Middle-Eastern. Drawing an association between geographical origin and appearance can prove difficult, especially since migration is associated with a geographical shift, and the transition introduces the opportunity for heterogeneity through intermarriage. It is evident that data surrounding the South Asian race is limited. Further, ethnicity verification for facial images of Pakistani origin, remains under-investigated with a single exception (Jilani et.al [14]).

IV. THE PAKISTANI FACE DATABASE

There are relatively few databases of facial images which are based on demographic information, such as race and ethnicity (Table I). Further, due to the lack of a standardized image capture process, there is considerable variability between published databases of facial images.

The Pakistani Face Database is a face image database, which has been developed by researchers within the University of Bradford, United Kingdom. A total of 463 students (280 male and 183 female) from the University of Bradford consented to have their photograph taken. However, of the total recruited participants 167 participants consented to only providing a frontal photograph. In total, 5 photographs were taken per subject using the HALO system: a custody image capture system [26].

The database is made up of 224 Pakistani and 239 Non-Pakistani participants. The ethnicity of each the image was determined by asking the subject to identify their ethnicity. For participants of Pakistani origin, eligibility was dependant on whether both the maternal and paternal parents were of Pakistani ethnicity.

TABLE I. FACIAL IMAGE DATASETS FOR FACIAL RECOGNITION

Database Name	**Published Facial Image Databases**		
	Target Race / Ethnicity	***Total Images***	***Author & Year***
Facial Recognition Technology Database (FERET)	Heterogeneous.	14,126	Phillips et al., 1998 [27]
NimStin Database	African American, Asian, European and Latino-American.	672	Tottenham et al., 2009 [28]
Chinese Emotional Expression	All Chinese.	75	Wang and Markham 1999 [29]
CAS-PEAL	All Chinese.	99,594	Gao et al., 2008 [30]
Asian Face Image Database PF01	Bengladeshi, Chinese, Korean and Vietnamese.	1,751	Dong and Gu 2001[31]
FACES	All Caucasian.	171	Ebner et al., 2010 [32]
Japanese Female Facial Expression (JAFFE) Database	All Japanese.	213	Lyons et al., 2014 [33]
Iranian Face Database (IFDB)	All Iranian.	Over 3,600	Bastanfard et al., 2007 [34]
Indian Movie Face Database (IMFD)	All South Asian (Indian).	34,512	Saha et al., 2012 [35]
Hajj and Umrah Dataset (HUFRD)	Unknown.	Unknown	Aly, 2012 [36]

Database Name	Published Facial Image Databases		
	Target Race / Ethnicity	*Total Images*	*Author & Year*
FEI Face Database	All Brazilian.	2,800	Thomaz and Giraldi 2010 [37]
The MR2	Africa, Asian and European.	74	Strohminger et al.,2016 [38]

A. Image Collection

Participants were requested to remove any items of clothing which obstructed the view of their face including their neck. Any hooded item was either removed or lowered. Those who wore a headscarf were requested to attend their session wearing a dark, non-patterned headscarf. Further, it was requested that make-up be kept to a minimum and participants who wore glasses were requested to remove them during image capture.

Participants were requested to not wear any jewellery/accessories from the collarbone above during their session. Participants were requested to sweep away long hair from their faces to ensure the periocular region was visible. During the process of photography, participants were requested to keep a neutral facial expression and to look directly into the camera ahead. A foot marker was mapped out onto the floor to ensure the participants stood at the correct distance from the central camera panel.

B. Image Processing

All the facial images went through a rigorous 'clean up' process. All the images were cropped around the face and neck, and effort was made to reduce any stray hair and remove earrings, if they were visible. Participants who wore a headscarf also went through the process of cropping, whereby the image was cropped in line with the drape of the worn headscarf, see figure 1. Facial blemishes such as mole or acne for example were also removed to prevent them being used as identification cues for future experiments.

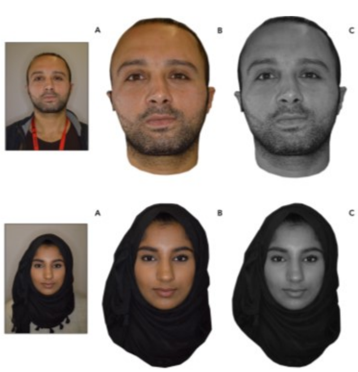

Figure 1. Example of face images that have been processed using Adobe Photoshop CS6, **(A)** Raw image captured using Halo, **(B)** cropped colour image, **(C)** greyscale image.

The images used have been reproduced in line with the approved ethics from the University of Bradford, UK and participant consent.

V. Methodology

Our approach for the Machine Learning experiments is to extract face features using three pre-trained ResNet models; 50 (50 layers), 101 (101 layers), and 152 (152 layers), and then to perform a binary classification using a Linear Support Vector Machine. The ResNet models were developed by He et al., [18] and are more accurate than a plain 34-layer Net. The algorithms are based on the premise of deep residual learning, where each residual block consisting of 3 layers' functions to fine-tune the previous blocks output by adding a learned residual to the input.

Each of the ResNet models were used in isolation of one another for each experiment. Experiments were conducted with the following datasets: (1) 1,000 full-face images, (2) 1,000 isolated eye crops, (3) isolated nose crops and (4) isolated mouth crops.

A. Data Pre-Processing

All images used in our experiments were resized to 224 × 224 pixels to ensure they conform to ResNet's input criteria. Moreover, data augmentation was carried out for the training data using rotations at the degree of 90°, 180° and 270° as well as random crops. This led to a two-fold increase in the original data size.

B. Feature Extraction

Neural Networks extract low-level face data such as skin colour and image edge information in addition to high-level face-shape data from localized features such as the eyes, nose and mouth. For feature representation, the activation of the last pooling layer of each of the ResNet models was used.

A total of three sets of features were retrieved using ResNet-50, RestNet-101, and ResNet-152 Neural Networks. We decided to exclude the last layer (output) of the Fully Connected layer (FC), since it was trained on a set of different data (i.e. objects) compared to the facial images which are presently used. Moreover, research [39] has demonstrated that the lower layers of the Deep Neural Network is sufficient in learning generic features.

C. Classification

A linear classifier was employed for binary ethnicity classification. Support Vector Machine (SVMs) are supervised machine learning models that function to identify a hyperplane, which best classifies data points within a given data space. Previous studies have demonstrated that SVMs is powerful as a binary classifier and operates by defining an optimum separating hyperplane between two classes of data [40] [41]. In our work, the two classes are Pakistani and Non-Pakistani.

Given a training set,

$$\{(x_i,\ y_i)\}_{1\le i\le n},\ x_i \in\ \mathbb{R}^d,\ y_i\ \in \{+1,\ -1\}, \tag{1}$$

SVM finds the Optimum Separating Hyperplane (OSH) by solving,

$$\begin{cases} \min\limits_{w,b} & \frac{1}{2} \parallel w \parallel^2 \\ \text{with} & y_i.(w.x+b) \ge 1 \end{cases} \tag{2}$$

for $i = 1\ \dots n$ are the observations, where the weights w and b the bias is learned during training. The classifier is learned such that, $y_i\ \in\ \{+1,\ -1\}$, $+1$ denotes facial images labelled as Pakistani and -1 denotes Non-Pakistani images. Thus, using a Support Vector Machine algorithm, an optimum separating hyperplane was computed, which separated the classes into the two categories [14]. A k-fold cross validation technique was employed to evaluate the performance of the pre-trained models.

VI. EXPERIMENTAL RESULTS

By extracting features from images using the pre-trained ResNet-50, ResNet-101, and ResNet-152 models, we achieved close to near perfect results for the binary classification of a Pakistani face using frontal image, (experiment 1) see table II.

TABLE II. PERFORMANCE ACCURACY FOR BINARY ETHNICITY VERIFICATION

Feature Extraction Model	**Classifier**
	Linear Support Vector Machine (SVM)
ResNet-50	98.8%
ResNet-101	***99.2%***
ResNet-152	99.0%

Having obtained the classification performance of the three pre-trained models on a dataset of 1,000 images, we conducted experiments to ascertain whether isolated features i.e. eyes, nose and mouth are also as informative and discriminatory for binary ethnicity classification. The additional 3 experiments were carried out using datasets of 1,000 eyes, nose and mouth. (Table III).

TABLE III. PERFORMANCE ACCURACY FOR ISOLATED FACE FEATURES FOR ETHNICITY VERIFICATION

Feature Extraction Model	**Face Feature: Eyes**
	Linear Support Vector Machine (SVM)
ResNet-50	85.9%
ResNet-101	85.6%
ResNet-152	***87.4%***

Feature Extraction Model	**Face Feature: Nose**
	Linear Support Vector Machine (SVM)
ResNet-50	91.5%
ResNet-101	***91.8%***
ResNet-152	91.7%

Feature Extraction Model	**Face Feature: Mouth**
	Linear Support Vector Machine (SVM)
ResNet-50	94.8%
ResNet-101	94.3%
ResNet-152	***95.7%***

The results achieved from the isolated features demonstrate that the eyes, mouth and nose are a reliable feature to separate between the Pakistani and Non-Pakistani class. Our results are comparable to published literature by Lyle et al., [42] who extracted gender and ethnicity information from images of the periocular region. Greyscale pixel intensities and periocular texture was computed by Local Binary Patterns (LBP) and a support vector machine was used for binary (Asian and Non-Asian) classification. Using a dataset of 4,232 images the researchers reported ethnicity classification at 91% with 5-fold cross validation. Qiu et al., [43] performed a two-class ethnicity task based on global texture analysis between Asian and Non-Asian, iris images. The researchers reported an accuracy of 85.95%, further demonstrating that the eyes can be used as a feature to differentiate amongst ethnicities.

Our additional experiments using the isolated nose and the mouth crops also provide promising results. Using a dataset of 1,000 images each, a performance accuracy of above 90% is achieved. The results are in line with published literature which suggests the eyes, nose and mouth are critical in determining ethnicity [2].

To determine the effectiveness of the results achieved, performance metrics such as sensitivity and specificity were calculated. Sensitivity is the value of positive cases classified correctly (TPR), and specificity is the value of negative cases classified correctly (TNR) [44]. By combining the two metrics the overall performance accuracy of the classification algorithm is reported. A visual representation of the values is shown in form of a Receiver Operating Characteristics (ROC) Graph. Figure 2 shows the

ROC for the full-face experiment using the three deep learning algorithms; ResNet-50, ResNet-101 and ResNet-152. All three models performed marginally close to perfect, which is depicted by the position of the performance lines, that are positioned to the left of the axis. However, it is difficult to clearly visualize the differences in accuracy, thus figure 3 best depicts the performance of ResNet-50, ResNet-101 and ResNet-152. Despite having seemingly lower accuracy, RestNet50, having an area under the curve (AUC) of 0.9990 has exhibited competitive performance. This proves that in the presence of sufficient face detail, the 50-layered architecture can capture strong ethnicity traits.

Figure 4 shows a column graph for the isolated mouth crops, which concluded the second highest results as per our experiments. While all three of the pre-trained models perform comparably, ResNet-152 outclassed Resnet-50 and ResNet-101 slightly. Thus, it is obvious that extremely deep neural networks extract relevant information, especially when there is very little facial detail.

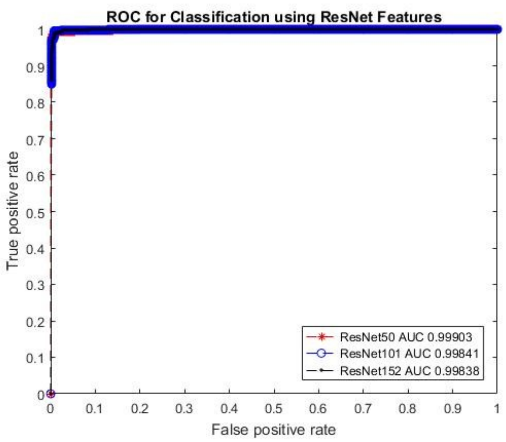

Figure 2. Receiver Operating Characteristics (ROC) Curve for the binary classification of the Pakistani full-face.

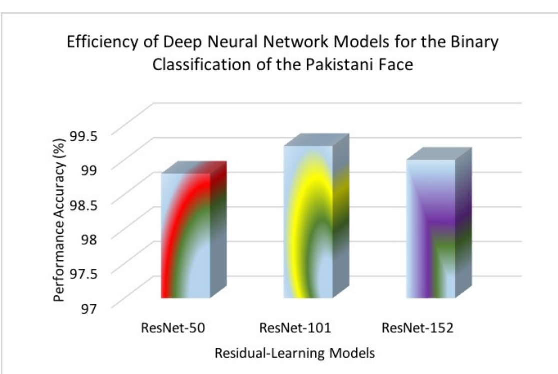

Figure 3. A column graph to illustrate the performance accuracy of ResNet-50, ResNet-101 and ResNet-152 for the binary classification of the Pakistani Face.

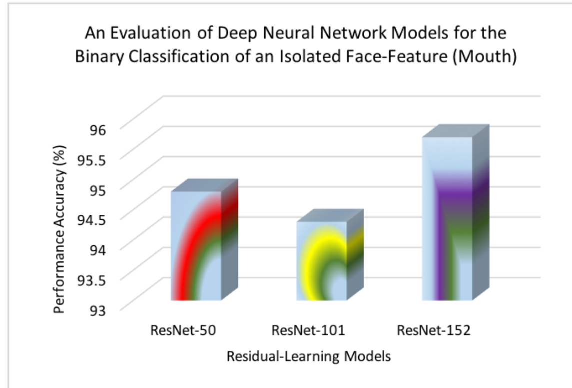

Figure 4. A column graph for the binary classification of an isolated face feature crop (mouth) using ResNet-50, ResNet-101 and ResNet-152.

VII. Discussion

The classification of human faces and its internal components i.e. eyes, nose and mouth, is an interesting challenge within the field of machine learning. The results presented within this paper are novel because to the best our knowledge, the application of deep learning algorithms for the ethnic classification of Pakistani faces and facial features remains unchallenged. The classification result displayed as part of table I shows as close to perfect performance accuracy for full-face images using ResNet-101 (99.2%). However, ResNet-50 and ResNet-152 performed equally as good with a performance accuracy of 98.8% and 99% respectively. The closeness of performance between the deep learning models is not surprising as He et al., [18] reported similar error rates for the models, when they were tested on the imageNet database.

The results we have achieved for the full-face ethnicity experiment using Residual Networks (ResNet), outperforms Ou et. al, [18] study on the binary classification and Hosoi et. al [2] study of a three-class ethnicity task. Since we achieved above 90% during the full face and isolated features experiment, namely for the nose and mouth.

It is fair to suggest that our results are in line with those reported by Muhammed et. al [23] who achieved 98.42%, 95.56%, 93.65%, 100% and 98.18% for the Asian, Black, Hispanic, Middle-Eastern and the White racial class respectively, using Local Binary Patterns (LBP). However, a distinction between our experimental results and those reported by other researchers is the difference in the dataset and as well as the proposed methodology. There is not an abundance of directly comparable literature on the classification of ethnicity from isolated facial features, yet our results are in-line with those reported by Momin et al., (2016) [45]. The authors used a fusion of 3 datasets to conduct multi-ethnic classifications (Asian Vs Non-Asian, Asian Vs White, Black and Indian). Classifiers such as k-means, Naïve Bayesian, Multilayer Perceptron (MLP) and Support Vector Machine (SVM) were used and ethnicity was tested on; right eye, left eye, nose and mouth.

Though the results for the binary classification for Asian features was consistently above 90%, the performance accuracy for the Black and Indian ethnic group were not as significant. The classification accuracy for the Indian nose on average (across all 4 classification techniques) was 49%, whereas the average for the mouth was 62%. Both the results are significantly below the performance accuracy we report as part of this paper. From all the experiments we have presented, the classification of the eyes showed a decline in classification. It can be suggested that it is not as informative as the nose and the mouth, when classifying the Pakistani ethnicity. Nevertheless, the results are promising and not futile since they can be used as a benchmark for further experiments if using either multiple classes or a different framework for feature extraction and classification.

It is important to mention that since the features of the lower third of the face (i.e. nose and mouth) gave a higher percentage of accuracy, this does not equate to superiority. Further experiments will be required to confirm whether the nose and mouth are more discriminatory compared to the eyes, if tested on other, varying databases. Furthermore, it is worth noting that for each feature extraction model used, there is the opportunity of training and testing each model with a different set of data. Currently, as per the experiments presented within this paper, the data is pooled from the same group of images. Ultimately, since there is no directly comparable literature on the South Asian, Pakistani ethnic group, it is hoped that the research presented in this paper will be used as a benchmark for future studies. The results achieved from all 4 sets of experiments demonstrate that deep learning algorithms can effectively learn and classify faces and features of Pakistani ethnicity. Moreover, given the diversity in facial appearance, faces which humans may struggle to ethnically classify, can accurately be discriminated by deep learning methods.

VIII. Conclusion

This paper addressed a two-class ethnicity classification problem based on facial images. Deep Learning models namely ResNet-50, ResNet-101 and ResNet-152 were used for the task of feature extraction which were then forwarded to a linear classifier. Results show the strength of performance of Deep Learning algorithms since our proposed framework concludes 99.2% accuracy for the ethnic classification of the Pakistani face using ResNet-101. A performance accuracy of 87.5%, 91.7% and 95.7% was achieved on a dataset of isolate eyes, nose and mouth crops respectively. In future work, we propose to use VGG-based models such as VGG-face, VGG-16 and VGG-19, in addition to a different classifier such as Decision Trees (DT). Additionally, we propose to increase the current dataset and include other regions of the face.

References

[1] Carcagnì, P., Del Coco, M., Cazzato, D., Leo, M. and Distante, C., 2015. A study on different experimental configurations for age, race, and gender estimation problems. *EURASIP Journal on Image and Video Processing*, *2015*(1), p.37.

[2] Hosoi, S., Takikawa, E. and Kawade, M., 2004, May. Ethnicity estimation with facial images. In Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. (pp. 195-200). IEEE.

[3] Hwang, H.-S., W.-S. Kim, and J.A. McNamara Jr, *Ethnic differences in the soft tissue profile of Korean and European-American adults with normal occlusions and well-balanced faces.* The Angle Orthodontist, 2002. **72**(1): p. 72-80.

[4] Pariona, A. (2017) *Ethnic Groups In Bangladesh.* Worldatlas.com. https://www.worldatlas.com/articles/ethnic-groups-in-bangladesh.html Accessed 24th April.

[5] International, M.R.G., *Redressing A History of Neglect: Discrimination of Ethnic Groups and Indigenous People of Pakistan*. 2009. p. 1-19 https://www.ecoi.net/en/file/local/1318355/1930_1334067966_mrgipakistan74.pdf Accessed 29th April.

[6] Hurst, C.O., *Pakistan's ethnic divide.* Studies in Conflict & Terrorism, 1996. **19**(2): p. 179-198.

[7] Fu, S., H. He, and Z.-G. Hou, *Learning race from face: A survey.* IEEE transactions on pattern analysis and machine intelligence, 2014. **36**(12): p. 2483-2509.

[8] LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436.

[9] Zhang, K., Huang, Y., Wu, H. and Wang, L., 2015, November. Facial smile detection based on deep learning features. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)* (pp. 534-538). IEEE.

[10] Xiao, T., Li, S., Wang, B., Lin, L. and Wang, X., 2017. Joint detection and identification feature learning for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3415-3424).

[11] Gao, W., et al., *The CAS-PEAL large-scale Chinese face database and baseline evaluations.* IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2008. **38**(1): p. 149-161

[12] Bastanfard, A., M.A. Nik, and M.M. Dehshibi, *Iranian face database with age, pose and expression.* Machine Vision, 2007: p. 50-55.

[13] Hwang, B.W., Roh, M.C. and Lee, S.W., 2004, May. Performance evaluation of face recognition algorithms on asian face database. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.* (pp. 278-283). IEEE.

[14] Jilani, S.K., Ugail, H., Bukar, A.M., Logan, A. and Munshi, T., 2017, September. A Machine Learning Approach for Ethnic Classification: The British Pakistani Face. In *2017 International Conference on Cyberworlds (CW)* (pp. 170-173). IEEE.

[15] Farkas, L. G., Katic, M. J. and Forrest, C. R. (2005) International anthropometric study of facial morphology in various ethnic groups/races. *Journal of Craniofacial Surgery* 16 (4), 615-646.

[16] Fang, F., Clapham, P. J. and Chung, K. C. (2011) A systematic review of inter-ethnic variability in facial dimensions. *Plastic and reconstructive surgery* 127 (2), 874.

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[18] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[19] Ou, Y., Wu, X., Qian, H. and Xu, Y., 2005, June. A real time race classification system. In *2005 IEEE International Conference on Information Acquisition* (pp. 6-pp). IEEE.

[20] Ahmed, A., Yu, K., Xu, W., Gong, Y. and Xing, E., 2008, October. Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks. In *European Conference on Computer Vision* (pp. 69-82). Springer, Berlin, Heidelberg.

[21] Guo, G. and Mu, G., 2010, June. A study of large-scale ethnicity estimation with gender and age variations. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops* (pp. 79-86). IEEE.

[22] Han, H., et al., *Demographic estimation from face images: Human vs. machine performance.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015. **37**(6): p. 1148-1161.

[23] Muhammad, G., Hussain, M., Alenezy, F., Bebis, G., Mirza, A. M. and Aboalsamh, H. (2012) Race classification from face images using local descriptors. *International Journal on Artificial Intelligence Tools* 21 (05), 1250019.

[24] Tin, H.H.K. and M.M. Sein, *Race identification for face images.* ACEEE Int. J. Inform. Tech, 2011. **1**(02): p. 35-37.

[25] Lakshmiprabha, N.S., Face Image Analysis using AAM, Gabor, LBP and WD features for Gender, Age, Expression and Ethnicity Classification. arXiv preprint arXiv:1604.01684, 2016.

[26] Jilani SK, Ugail H, Cole S and Logan A (2018) Standardising the Capture and Processing of Custody Images. Current Journal of Applied Science and Technology. 30(5): 1-13.

[27] Phillips, P. J., Wechsler, H., Huang, J. and Rauss, P. J. (1998) The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing* 16 (5), 295-306.

[28] Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J. and Nelson, C. (2009) The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research* 168 (3), 242-249.

[29] Wang, L. and Markham, R. (1999) The development of a series of photographs of Chinese facial expressions of emotion. *Journal of Cross-Cultural Psychology* 30 (4), 397-410.

[30] Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X. and Zhao, D. (2008) The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 38 (1), 149-161.

[31] Dong, H. and Gu, N. (2001) Asian face image database PF01.

[32] Ebner, N. C., Riediger, M. and Lindenberger, U. (2010) FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior research methods* 42 (1), 351-362.

[33] Lyons, M. J., Kamachi, M. and Gyoba, J. (2014) Japanese Female Facial Expressions (JAFFE), Database of digital images. 2007.

[34] Bastanfard, A., Nik, M. A. and Dehshibi, M. M. (2007) Iranian face database with age, pose and expression. *Machine Vision*, 50-55.

[35] Saha, K., Debnath, R., Bhowmik, M.K., Bhattacharjee, D. and Nasipuri, M., 2012. North-east Indian face database: its design and aspects.

[36] Aly, S. A. (2012) Pilgrims Face Recognition Dataset--HUFRD. *arXiv preprint arXiv:1205.4463.*

[37] Thomaz, C. E. and Giraldi, G. A. (2010) A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing* 28 (6), 902-913.

[38] Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C. and Heagins, T. B. (2016) The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior research methods* 48 (3), 1197-1204.

[39] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L., 2014. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 1725-1732).

[40] Chierchia, G., Pustelnik, N. and Pesquet, J.C., 2016, September. Random primal-dual proximal iterations for sparse multiclass SVM. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE.

[41] Nalavade, K. and Meshram, B.B., 2012. Data Classification Using Support Vector Machine. In *National Conference on Emerging Trends in Engineering & Technology (VNCET)* (Vol. 2, pp. 181-184).

[42] Lyle, J.R., Miller, P.E., Pundlik, S.J. and Woodard, D.L., 2010, September. Soft biometric classification using periocular region features. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (pp. 1-7). IEEE.

[43] Qiu, X., Sun, Z. and Tan, T., 2006, January. Global texture analysis of iris images for ethnic classification. In *International Conference on Biometrics* (pp. 411-418). Springer, Berlin, Heidelberg.

[44] Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation

[45] Momin, Hajra & Tapamo, Jules-Raymond. (2016). A Comparative study of a Face Components based Model of Ethnic Classification using Gabor Filters. Applied Mathematics & Information Sciences. 10. 2255-2265. 10.18576/amis/10062.